IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Applicant:	Kas Kasravi et al.	Examiner: Leonard Saint Cyr
Serial No.:	10/766,308	Group Art Unit: 2626 Conf. 1219
Filed:	January 27, 2004	Docket No.: 200901335-1
Title:	SYSTEM AND METHOD FOR COMPARATIVE ANALYSIS OF TEXTUAL DOCUMENTS	

APPEAL BRIEF UNDER 37 C.F.R. §41.37

Mail Stop Appeal Brief – Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir/Madam:

This Appeal Brief is submitted in support of the Amendment and Response and concurrently filed Notice of Appeal filed on November 11, 2009, appealing the final rejection of claims 1-3, 6, 7, 10-14, 16-21, 23, 24, 26-31, 36-and 40 of the above-identified application as set forth in the Final Office Action mailed August 10, 2009, and after the Amendment After Final indicated as entered on February 12, 2010.

The U.S. Patent and Trademark Office is hereby authorized to charge Deposit Account No. 50-0471 in the amount of $540.00 for filing a Brief in Support of an Appeal as set forth under 37 C.F.R. § 41.20(b)(2). At any time during the pendency of this application, please charge any required fees or credit any overpayment to Deposit Account No. 50-0471.

TABLE OF CONTENTS

REAL PARTY IN INTEREST

The real party in interest is Hewlett-Packard Development Company, LP having a principal place of business at 11445 Compaq Center Drive West, Houston, TX 77070, U.S.A. (hereinafter "HPDC"). HPDC is a Texas limited partnership and is a wholly-owned affiliate of Hewlett-Packard Company, a Delaware corporation, headquartered in Palo Alto, CA. The general or managing partner of HPDC is HPQ Holdings, LLC.

RELATED APPEALS AND INTERFERENCES

There are no other appeals or interferences known to Appellant that will have a bearing on the Board's decision in the present Appeal.

STATUS OF CLAIMS

Claims 1-40 were pending in the application. Appellant had previously cancelled claims 4-5, 8-9, 15, 22, 25, and 32. The Final Office Action mailed August 10, 2009, allowed claims 33 and 35, and an amendment entered after final allowed claim 34. Claims 1-3, 6, 7, 10-14, 16-21, 23, 24, 26-31, and 36-40 remain rejected. These rejected claims are the subject of the present Appeal.

STATUS OF AMENDMENTS

An Amendment and Response was concurrently filed with the Notice of Appeal on November 10, 2009, subsequent to the Final Office Action mailed on August 10, 2009, which was indicated as entered on February 12, 2010, removing the rejection based on 35 U.S.C 101 and allowing claim 34.

SUMMARY OF THE CLAIMED SUBJECT MATTER

Claim 1 is directed to a computer-implemented method of comparing the semantic content of two or more documents (page 4, paragraph [00015], lines 1-4). The method includes accessing a plurality of documents (page 4, paragraph [00015], lines 1-2); performing a linguistic analysis on each document (page 10, paragraph [00032], lines 1-3); and defining a semantic vector for each document based on the linguistic analysis, where the semantic vector includes multiple components (page 12, paragraph [00036], lines 1-3; see also generally paragraph [00039]). Each component of said semantic vector has at least a term included in the document or a synonym of said term (page 13, paragraph [00039], lines 20-21); a weighting factor relating to an importance, based on characteristics of the document, of the term (page 14, paragraph [00040], lines 8-10); and a frequency value relating to a number of occurrences of said term (page 20, paragraph [00054], lines 1-2). The semantic vector is processed by a digital computer. The method further includes comparing a semantic vector of an identified document to the semantic vector for each document in the plurality of documents to determine at least one document semantically similar to the identified document (page 12, paragraph [00037], lines 1-4).

Claim 14 is directed to a computer-implemented method of comparing two or more documents (page 4, paragraph [00015], lines 1-4). The method includes linguistically analyzing a plurality of documents to identify at least one term group in each document, where each term group comprising a main term and at least one subordinate term semantically related to the main term (page 32, paragraph [00085], lines 5-7). The method also includes generating a semantic vector associated with each document, the semantic vector comprising a plurality of components (page 12, paragraph [00036], lines 1-3). Each component includes a term group in the document (page 32, paragraph [00085], line 5); a frequency value relating to a number of occurrences of the term group (page 32, paragraph [00085], line 6); and a weighting factor relating to an importance, based on characteristics of the document, of at least part of the term group (page 32, paragraph [00085], line 6-7). The semantic vector is processed by a digital computer. The method also includes comparing a semantic vector of an identified document to the semantic vector for each document in the plurality of documents to determine at least one document semantically similar to the

identified document using a defined metric (page 12, paragraph [00037], lines 1-4). The metric measures the semantic distance between documents as a function of at least the frequency values included in the semantic vectors for the documents (page 33, paragraph [00087], lines 6-7).

Claim 23 is directed to a system for comparing two or more documents (page 4, paragraph [00015], lines 1-4). The system includes a document inputter, a semantic analyzer, a semantic quantifier, and a comparator. The document inputter is arranged to access a plurality of documents (page 31, paragraph [00082], lines 1-7). The semantic analyzer is arranged to perform a linguistic analysis on each document to identify at least one term group in the document, each term group comprising a main term and at least one subordinate term semantically related to the main term (page 32, paragraph [00085], lines 5-7). The semantic quantifier is arranged to output a quantified representation of a semantic content of each document (page 33, paragraph [00087], lines 1-7). The quantified representation based at least in part on a term group in the document and a weighting factor relating to an importance, based on characteristics of the document, of at least part of the term group (page 33, paragraph [00086], lines 1-7). The comparator is arranged to compare the quantified representations using a defined algorithm, where the defined algorithm measures the semantic distance between documents as a function of at least the weighting factors associated with the quantified representations for the documents (page 36, paragraph [00089], lines 1-7) to determine at least one document in the plurality of documents semantically similar to an identified document (page 12, paragraph [00037], lines 1-4).

Claim 24 is directed to a system for comparing two or more documents (page 4, paragraph [00015], lines 1-4). The system includes a document inputter, a semantic analyzer, a semantic vector generator, and a comparator. The document inputter is arranged to access a plurality of documents (page 31, paragraph [00082], lines 1-7). The semantic analyzer is arranged to perform a linguistic analysis on each document to identify at least one term group in the document, each term group comprising a main term and at least one subordinate term semantically related to the main term (page 32, paragraph [00085], lines 5-7). The semantic vector generator is arranged to output a semantic vector associated with each document (page 33, paragraph [00087], lines 1-7). Each semantic vector includes a plurality of components, and each component includes a term group in the document (page 32, paragraph [00085], line

5); a frequency value relating to a number of occurrences of the term group (page 32, paragraph [00085], line 6); and a weighting factor relating to an importance, based on characteristics of the document, of at least part of the term group (page 32, paragraph [00085], lines 6-7). The comparator is arranged to compare the semantic vectors using a defined metric, wherein said metric measures the semantic distance between documents as a function of at least the frequency values included in the semantic vectors for the documents (page 33, paragraph [00087], lines 1-7) to determine at least one document in the plurality of documents semantically similar to an identified document (page 12, paragraph [00037], lines 1-4).

Claim 26 is directed to a computer program product comprising a computer usable medium having computer readable program code means embodied therein (page 4, paragraph [00015], lines 1-4). The computer readable program code means in said computer program product includes means for causing a computer to access a plurality of documents (page 31, paragraph [00082], lines 1-7); to perform a linguistic analysis on each document to identify at least one term group in the document, where each term group comprising a main term and at least one subordinate term semantically related to the main term (page 32, paragraph [00085], lines 5-7); and output a quantified representation of a semantic content of each document (page 33, paragraph [00087], lines 1-7). The quantified representation based at least in part on: a term group in the document, a frequency value relating to a number of occurrences of the term group, and a weighting factor relating to an importance, based on characteristics of the document, of at least part of the term group (page 32, paragraph [00085], lines 5-7). The computer readable program code means in said computer program product also includes means for causing a computer to compare the quantified representations using a defined algorithm, where the defined metric measures the semantic distance between documents as a function of at least the frequency values associated with the quantified representations for the documents (page 33, paragraph [00087], lines 1-7) to determine at least one document in the plurality of documents semantically similar to an identified document (page 12, paragraph [00037], lines 1-4).

Claim 27 is directed to a computer program product comprising a computer usable medium having computer readable program code means embodied therein (page 4, paragraph [00015], lines 1-4). The computer readable program code means in said computer program

product includes means for causing a computer to linguistically analyze a plurality of documents to identify at least one term group in the document, each term group comprising a main term and at least one subordinate term semantically related to the main term (page 32, paragraph [00085], lines 5-7) and to generate a semantic vector associated with each document (page 12, paragraph [00036], lines 1-3). Each semantic vector comprising a plurality of components, each component includes a term group in the document and a weighting factor relating to an importance, based on characteristics of the document, of at least part of the term group (page 33, paragraph [00086], lines 1-7). The computer readable program code means in said computer program product also includes means for causing a computer to compare the semantic vectors using a defined metric, wherein said metric measures the semantic distance between semantic vectors as a function of at least the weighting factors included in the semantic vectors (page 36, paragraph [00089], lines 1-7) to determine at least one document in the plurality of documents semantically similar to an identified document (page 12, paragraph [00037], lines 1-4).

GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

I. Claims 1, 2, 3, 6, 7, 10-14, 16-21, 23, 24, 26-31, and 36-40 stand rejected under 35 U.S.C. 102(a) as being anticipated by Gillis (U.S. Patent 6,523,026).

ARGUMENT

I. The Applicable Law

To anticipate a claim under 35 U.S.C. 102, a reference must teach every limitation of the claim. *Verdegaal Bros. v. Union Oil Co. of California*, 814 F.2d 628, 631 [2 USPQ2d 1051, 1053] (Fed. Cir. 1987) ("A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference"). *See also Atlas Power Co. v. IRECO Inc.*, 190 F.3d 1342, 1347 [51 USPQ2d 1943, 1946] (Fed. Cir. 1999).

II. Rejection of Claims 1, 2, 3, 6, 7, 10-14, 16-21, 23, 24, 26-31, and 36-40 under 35 U.S.C. §102(a)

Claims 1, 2, 3, 6, 7, 10-14, 16-21, 23, 24, 26-31, and 36-40 stand rejected under 35 U.S.C. 102(a) as being anticipated by U.S. Patent No. 6,523,026 to Gillis (the Gillis Patent). Claims 1, 14, 23, 24, 26, 27 are independent claims. Claims 2, 3, 6, 7, 10-13, 36, and 37 depend from claim 1. Claims 16-21 and 38-40 depend from claim 14. Claims 28-31 depend from claim 27 No claims depend from claims 23, 24, and 26.

Independent claim 1 recites, "comparing a semantic vector of an identified document to the semantic vector for each document in the plurality of documents to determine at least one document semantically similar to the identified document." In particular, a semantic vector of a document is compared to a semantic vector for each document in a plurality of documents. Independent claims 14, 23, 24, 26, and 27 recite similar limitations. In contrast, the Gillis Patent merely teaches comparing term vectors to a target domain to identify semantically distant analogies.

The Final Office Action at page 2 states that this claimed feature is taught at column 7, lines 59-66 of the Gillis Patent. This teaching, made in describing a feature of latent

semantic indexing, or LSI, which is described in the Gillis patent as having "difficulty handling words with multiple, domain specific, meanings (polysemy)" at column 8, lines 18-19. The cited teaching of the Gillis patent is taken out of context, in that is describes a difficulty the disclosure of the Gillis Patent seeks to overcome. Accordingly, this cited feature is not to be included with the features of the Gillis patent cited against the claims. It is specifically excluded from the remaining features of the claims and it does not belong as part of combination of pending independent claim elements.

Accordingly, because the Gillis Patent does not teach every feature of the independent claims 1, 14, 23, 24, 26, and 27, the claims are patentably distinguishable from the Gillis Patent. The dependent claims serve to further define the subject matter of their respective independent claims and are also patentably by virtue of their dependency. Accordingly, Appellant request that the rejection of claims 1, 2, 3, 6, 7, 10-14, 16-21, 23, 24, 26-31, and 36-40 based on 35 U.S.C. 102(a) be reversed and the claims be allowed.

CONCLUSION

For the above reasons, Appellants respectfully submit that the cited references neither anticipate nor render obvious claims of the pending Application. The pending claims distinguish over the cited references, and therefore, Appellants respectfully submit that the rejections must be withdrawn, and respectfully request the Examiner be reversed and claims 1-3, 6, 7, 10-14, 16-21, 23, 24, 26-31, and 34-40 be allowed.

Any inquiry regarding this Response should be directed to Patrick G. Billig at Telephone No. (612) 573-2003, Facsimile No. (612) 573-2005.

Respectfully submitted,

Kas Kasravi et al.,

By their attorneys,

DICKE, BILLIG & CZAJA, PLLC
Fifth Street Towers, Suite 2250
100 South Fifth Street
Minneapolis, MN 55402
Telephone: (612) 573-2003
Facsimile: (612) 573-2005

Dated: February 12, 2010
PGB:RPH:cms

/Patrick G. Billig/
Patrick G. Billig
Reg. No. 38,080

CLAIMS APPENDIX

1. A computer-implemented method of comparing the semantic content of two or more documents, comprising:

accessing a plurality of documents;

performing a linguistic analysis on each document;

defining a semantic vector for each document based on the linguistic analysis, said semantic vector having multiple components, wherein each component of said semantic vector has at least:

a term included in the document or a synonym of said term;

a weighting factor relating to an importance, based on characteristics of the document, of said term; and

a frequency value relating to a number of occurrences of said term;

processing the semantic vector by a digital computer; and

comparing a semantic vector of an identified document to the semantic vector for each document in the plurality of documents to determine at least one document semantically similar to the identified document.

2. The method of claim 1, wherein the linguistic analysis comprises sentence analysis.

3. The method of claim 2, wherein the sentence analysis comprises a syntactic analysis and a semantic analysis.

4. (Canceled)

5. (Canceled)

6. The method of claim 1, wherein each component of the semantic vector for at least one of the documents comprises multiple dimensions.

7. The method of claim 1, wherein each component of the semantic vector for at least one of the documents further comprises a subordinate concept value.

8. (Canceled)

9. (Canceled)

10. The method of claim 1, wherein some of the components of the semantic vector for at least one of the documents have main term – subordinate term pairs as their first value.

11. The method of claim 1, wherein the semantic vector comprises a multi-dimensional vector defined by the content of a semantic net.

12. The method of claim 11, wherein the content of the semantic net is augmented by relative weights, strengths, or frequencies of occurrence of the features within the semantic net.

13. The method of claim 37, wherein an output of said defined algorithm is a measure of at least one of semantic distance, semantic similarity, semantic dissimilarity, degree of patentable novelty and degree of anticipation.

14. A computer-implemented method of comparing two or more documents, comprising:

linguistically analyzing a plurality of documents to identify at least one term group in each document, each term group comprising a main term and at least one subordinate term semantically related to the main term;

generating a semantic vector associated with each document, the semantic vector comprising a plurality of components, each component including:

a term group in the document;

a frequency value relating to a number of occurrences of the term group; and

a weighting factor relating to an importance, based on characteristics of the document, of at least part of the term group;

processing the semantic vector by a digital computer; and

comparing a semantic vector of an identified document to the semantic vector for each document in the plurality of documents to determine at least one document semantically similar to the identified document using a defined metric, wherein said metric measures the semantic distance between documents as a function of at least the frequency values included in the semantic vectors for the documents.

15. (Canceled)

16. The method of claim 14, wherein the main term includes synonyms of the main term.

17. The method of claim 14, wherein one or more of said two or more documents are located using an autonomous software or 'bot program.

18. The method of claim 17, wherein the 'bot program automatically analyzes each document in a defined domain or network by executing a series of rules and assigning an overall score to the document.

19. The method of claim 18, wherein all documents with a score above a defined threshold are linguistically analyzed.

20. The method of claim 14, wherein the semantic vector is a quantification of the semantic content of each document.

21. The method of claim 14, wherein each component has multiple dimensions.

22. (Canceled)

23. A system for comparing two or more documents, comprising:

a document inputter, arranged to access a plurality of documents;

a semantic analyzer, arranged to perform a linguistic analysis on each document to identify at least one term group in the document, each term group comprising a main term and at least one subordinate term semantically related to the main term;

a semantic quantifier, arranged to output a quantified representation of a semantic content of each document, the quantified representation based at least in part on:

a term group in the document; and

a weighting factor relating to an importance, based on characteristics of the document, of at least part of the term group; and

a comparator, arranged to compare the quantified representations using a defined algorithm, wherein said defined algorithm measures the semantic distance between documents as a function of at least the weighting factors associated with the quantified representations for the documents to determine at least one document in the plurality of documents semantically similar to an identified document.

24. A system for comparing two or more documents, comprising:

a document inputter, arranged to access a plurality of documents;

a semantic analyzer, arranged to perform a linguistic analysis on each document to identify at least one term group in the document, each term group comprising a main term and at least one subordinate term semantically related to the main term;

a semantic vector generator, arranged to output a semantic vector associated with each document, each semantic vector comprising a plurality of components, each component including:

a term group in the document;

a frequency value relating to a number of occurrences of the term group; and

a weighting factor relating to an importance, based on characteristics of the document, of at least part of the term group; and

a comparator, arranged to compare the semantic vectors using a defined metric, wherein said metric measures the semantic distance between documents as a function of at least the frequency values included in the semantic vectors for the documents to determine at

least one document in the plurality of documents semantically similar to an identified document.

25. (Canceled)

26. A computer program product comprising a computer usable medium having computer readable program code means embodied therein, the computer readable program code means in said computer program product comprising means for causing a computer to:

access a plurality of documents;

perform a linguistic analysis on each document to identify at least one term group in the document, each term group comprising a main term and at least one subordinate term semantically related to the main term;

output a quantified representation of a semantic content of each document, the quantified representation based at least in part on:

a term group in the document;

a frequency value relating to a number of occurrences of the term group; and

a weighting factor relating to an importance, based on characteristics of the document, of at least part of the term group; and

compare the quantified representations using a defined algorithm, wherein said defined metric measures the semantic distance between documents as a function of at least the frequency values associated with the quantified representations for the documents to determine at least one document in the plurality of documents semantically similar to an identified document.

27. A computer program product comprising a computer usable medium having computer readable program code means embodied therein, the computer readable program code means in said computer program product comprising means for causing a computer to:

linguistically analyze a plurality of documents to identify at least one term group in the document, each term group comprising a main term and at least one subordinate term semantically related to the main term;

generate a semantic vector associated with each document, each semantic vector

comprising a plurality of components, each component including:

a term group in the document; and

a weighting factor relating to an importance, based on characteristics of the document, of at least part of the term group; and

compare the semantic vectors using a defined metric, wherein said metric measures the semantic distance between semantic vectors as a function of at least the weighting factors included in the semantic vectors to determine at least one document in the plurality of documents semantically similar to an identified document.

28. The computer program product of claim 27, wherein the computer readable program code means in said computer program product further comprises means for causing a computer to:

identify one or more of said two or more documents using an autonomous software or 'bot program.

29. The computer program product of claim 28, wherein said 'bot program automatically analyzes each document in a defined domain or network by executing a series of rules and assigning an overall score to the document.

30. The computer program product of claim 27, wherein the semantic vector is a quantification of the semantic content of each document.

31. The computer program product of claim 27, wherein an output of said defined metric is a measure of at least one of semantic distance, semantic similarity, semantic dissimilarity, degree of patentable novelty and degree of anticipation.

32. (Canceled)

33. (Allowed) A system for comparing two or more documents, comprising:

a document inputter, arranged to access two or more documents;

a semantic analyzer, arranged to perform a linguistic analysis on each document;

a semantic vector generator, arranged to output a semantic vector associated with each document; and

a comparator, arranged to compare the semantic vectors using a defined metric, wherein said defined metric is one of:

$[Sqrt(f1^2 + f2^2 + f3^2 + f4^2 + + f(N-1)^2 fN^2)/n] * 100$, wherein f is a difference in frequency of a common term between two documents and n is the number of terms those documents have in common; or

Sqrt(sum((w-Delta)^2 * w-Avg))/(Log(n)^3 * 1000), wherein w-Delta is the difference in weight between two common terms, w-Avg is the average weight between two common terms, and n is the number of common terms, between two documents.

34. (Allowed) A computer-implemented method of comparing two or more documents, comprising:

linguistically analyzing a plurality of documents;

generating a semantic vector associated with each document;

processing the semantic vector by a digital computer; and

comparing the semantic vectors using a defined metric, wherein said defined metric is one of:

$[Sqrt(f1^2 + f2^2 + f3^2 + f4^2 + + f(N-1)^2 fN^2)/n] * 100$, wherein f is a difference in frequency of a common term between documents and n is the number of terms those documents have in common; or

Sqrt(sum((w-Delta)^2 * w-Avg))/(Log(n)^3 * 1000), wherein W-Delta is the difference in weight between two common terms, w-Avg is the average weight between two common terms, and n is the number of common terms, between documents to determine at least one document in the plurality of documents semantically similar to an identified document.

35. (Allowed) A computer program product comprising a computer usable medium having computer readable program code means embodied therein, the computer readable program code means in said computer program product comprising means for causing a computer to access two or more documents;

perform a linguistic analysis on each document;

output a quantified representation of a semantic content of each document; and

compare the quantified representations using a defined algorithm, wherein said defined algorithm is one of:

$[Sqrt(f1^2 + f2^2 + f3^2 + f4^2 + + f(N-1)^2 fN^2)/n] * 100$, wherein f is a difference in frequency of a common term between two documents and n is the number of terms those documents have in common; or

Sqrt(sum((w-Delta)^2 * w-Avg))/(Log(n)^3 * 1000), wherein w-Delta is the difference in weight between two common terms, w-Avg is the average weight between two common terms, and n is the number of common terms, between two documents.

36. The method of claim 1, wherein said term comprises at least one of a word or a phrase.

37. The method of claim 1, further comprising comparing the semantic vectors based on a defined algorithm.

38. The method of claim 14, wherein the at least one subordinate term includes synonyms of one of the subordinate terms.

39. The method of claim 14, wherein one or more of the at least one subordinate term or the main term comprises a phrase.

40. The method of claim 14, wherein the weighting factor comprises a plurality of different weighting factors and each of the different weighting factors relates to the importance of the main term or a subordinate term in the term group.

EVIDENCE APPENDIX

None.

RELATED PROCEEDINGS APPENDIX

None.